

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

894936271
CA 00258

CA00258

27

JU

2000 (27.06.00)

REC'D	11 JUL 2000
WIPO	PCT

PA 232820



THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

April 10, 2000

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: 60/124,260

FILING DATE: March 11, 1999

PRIORITY
DOCUMENT

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS



P. SWAIN
Certifying Officer

66/TT/99
JC375 U.S. PTO

Appro
JC541 U.S. PTO
66/124266
6/1/99

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

DOCKET NUMBER

11757.30USP1

CERTIFICATE UNDER 37 CFR 1.10:
"Express Mail" mailing label number: EL119854905US
Date of Deposit: March 11, 1999

I hereby certify that this paper or fee is being deposited with the U.S. Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to Assistant Commissioner for Patents, Washington, D.C. 20231.

By: *Walter White*
Name: Walter White

REQUEST FOR PROVISIONAL APPLICATION UNDER 37 C.F.R. § 1.53(c)

BOX PROVISIONAL PATENT APPLICATION
Assistant Commissioner for Patents
Washington, DC 20231

Dear Sir:

This is a request for filing a Provisional application for patent under 37 CFR § 1.53(c) entitled NOVEL KALLIKREIN-LIKE GENES by the following inventor(s):

Full Name Of Inventor	Family Name Diamandis	First Given Name Eleftherios	Second Given Name P.
Residence & Citizenship	City Toronto	State or Foreign Country Ontario, Canada	Country of Citizenship Canada
Post Office Address	Post Office Address 44 Gerrard St. West, Suite 1504	City Toronto	State & Zip Code/Country Ontario Canada M5G 2K2
Full Name Of Inventor	Family Name	First Given Name	Second Given Name
Residence & Citizenship	City	State or Foreign Country	Country of Citizenship
Post Office Address	Post Office Address	City	State & Zip Code/Country
Full Name Of Inventor	Family Name	First Given Name	Second Given Name
Residence & Citizenship	City	State or Foreign Country	Country of Citizenship
Post Office Address	Post Office Address	City	State & Zip Code/Country

- Enclosed is the Provisional application for patent as follows: 44 pages of specification, and 1 sheets of drawings.
- A Verified Statement that this filing is by a small entity (37 CFR 1.9, 1.27, 1.28) is attached.
- Payment of Provisional filing fee under 37 C.F.R. § 1.16(k):
 Attached is a check in the amount of \$ 75.00.
 Please charge Deposit Account No. 13-2725.
 PAYMENT OF THE FILING FEE IS BEING DEFERRED.
- The Commissioner is hereby authorized to charge any additional fees as set forth in 37 CFR §§ 1.16 to 1.18 which may be required by this paper or credit any overpayment to Account No. 13-2725.

5. Enclosed is an Assignment of the invention to , Recordation Form Cover Sheet and a check for \$ to cover the Recordation Fee.

6. Also Enclosed:

7. The invention was made by the following agency of the United States Government or under a contract with the following agency of the United States Government:

8. Address all future communications to the **Attention of Douglas P. Mueller** (may only be completed by attorney or agent of record) at the address below.

9. A return postcard is enclosed.

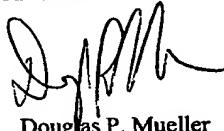
Respectfully submitted,

Eleftherios P. Diamandis

By his Attorneys,

MERCHANT, GOULD, SMITH, EDELL,
WELTER & SCHMIDT, P.A.
3100 Norwest Center
90 South Seventh Street
Minneapolis, Minnesota 55402
(612) 332-5300

By



Douglas P. Mueller
Reg No. 30,300
DPM:vvh

Dated: March 11, 1999

60124660 031199

MSH File: KALLIKREIN

UNITED STATES PROVISIONAL

6674100 - 09/24/2010

Title: Novel Human Kallikrein-Like Genes~~

Inventors: Eleftherios P. Diamandis

MSH File : KALLIKREIN

TITLE: Novel Human Kallikrein-Like Genes

FIELD OF THE INVENTION

The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid
5 molecules; and use of the proteins and nucleic acid molecules

BACKGROUND OF THE INVENTION

Kallikreins and kallikrein-like proteins are a subgroup of the serine protease enzyme family and exhibit a high degree of substrate specificity (1). The biological role of these kallikreins is the selective cleavage of specific polypeptide precursors (substrates) to release 10 peptides with potent biological activity (2). In mouse and rat, kallikreins are encoded by large multigene families. In the mouse genome, at least 24 genes have been identified (3). Expression of 11 of these genes has been confirmed; the rest are presumed to be pseudogenes (4). A similar family of 15-20 kallikreins has been found in the rat genome (5) where at least 4 of these are known to be expressed (6).

15 Three human kallikrein genes have been described, i.e. prostatic specific antigen (PSA or KLK3) (7), human glandular kallikrein (KLK2) (8) and tissue (pancreatic-renal) kallikrein (KLK1) (9). The PSA gene spans 5.8 Kb of sequence which has been published (7); the KLK2 gene has a size of 5.2 Kb and its complete structure has also been elucidated (8). The KLK1 gene is approximately 4.5 Kb long and the exon sequences and the exon/intron junctions of this 20 gene have been determined (9).

The mouse kallikrein genes are clustered in groups of up to 11 genes on chromosome 7 and the distance between the genes in the various clusters can be as small as 3-7 Kb (3). All three human kallikrein genes have been assigned to chromosome 19q13.2 – 19q13.4 and the distance between PSA and KLK2 has been estimated to be 12 Kb (9).

25 A major difference between mouse and human kallikreins is that two of the human kallikreins (KLK2 and KLK3) are expressed almost exclusively in the prostate while in animals none of the kallikreins is localized in this organ. Other candidate new members of the human kallikrein gene family include protease M (10) (also named Zyme (11) or neurosin (12) and the normal epithelial cell-specific gene-1 (NES1) (13). Both genes have been assigned to 30 chromosome 19q13.3 (10,14) and show structural homology with other serine proteases and the kallikrein gene family (10-14).

SUMMARY OF THE INVENTION

In efforts to precisely define the relative genomic location of PSA, KLK2, Zyme and NES1 genes, an area spanning approximately 300 Kb of contiguous sequence on human chromosome 19 (19q13.3 -q13.4) was examined. The present inventor was able to identify the 5 relative location of the known kallikrein genes and, in addition, he identified other putative kallikrein- like genes which exhibit both location proximity and structural similarity with the known members of the human kallikrein family. The novel genes exhibit homology with the currently known members of the kallikrein family and they are co-localized in the same genomic region. These new genes, like the already known kallikreins have utility in various cancers 10 including those of the breast and prostate.

The kallikrein-like proteins described herein are individually referred to as "KLK-L1 to KLK-L8", and collectively as "kallikrein-like proteins" or "KLK-L Proteins". The genes encoding the proteins are referred to as *klk-l1* to *klk-l8* or kallikrein-like genes or "*klk-l* genes".

Broadly stated the present invention relates to an isolated nucleic acid molecule which 15 comprises:

- (i) a nucleic acid sequence encoding a protein having substantial sequence identity preferably at least 70% sequence identity, with an amino acid sequence of KLK-L1-KLK-L8 as shown in Tables 2 to 9; or
- (ii) a nucleic acid sequence encoding a protein comprising with an amino acid sequence of KLK-L1-KLK-L8 as shown in Tables 2 to 9;
- (iii) nucleic acid sequences complementary to (i);
- (iv) a degenerate form of a nucleic acid sequence of (i);
- (v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);
- 20 (vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising with an amino acid sequence of KLK-L1-KLK-L8 as shown in Tables 2 to 9; or
- (vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

The invention also contemplates a nucleic acid molecule comprising a sequence 30 encoding a truncation of a KLK-L protein, an analog, or a homolog of a KLK-L Protein or a truncation thereof. (KLK-L Protein and truncations, analogs and homologs of the KLK-L

Protein are also collectively referred to herein as "KLK-L Related Proteins").

The nucleic acid molecules of the invention may be inserted into an appropriate expression vector, i.e. a vector that contains the necessary elements for the transcription and translation of the inserted coding sequence. Accordingly, recombinant expression vectors 5 adapted for transformation of a host cell may be constructed which comprise a nucleic acid molecule of the invention and one or more transcription and translation elements linked to the nucleic acid molecule.

The recombinant expression vector can be used to prepare transformed host cells expressing KLK-L Related Proteins. Therefore, the invention further provides host cells 10 containing a recombinant molecule of the invention. The invention also contemplates transgenic non-human mammals whose germ cells and somatic cells contain a recombinant molecule comprising a nucleic acid molecule of the invention, in particular one which encodes an analog of the KLK-L Protein, or a truncation of the KLK-L Protein.

The invention further provides a method for preparing KLK-L Related Proteins utilizing 15 the purified and isolated nucleic acid molecules of the invention. In an embodiment a method for preparing a KLK-L Related Protein is provided comprising (a) transferring a recombinant expression vector of the invention into a host cell; (b) selecting transformed host cells from untransformed host cells; (c) culturing a selected transformed host cell under conditions which allow expression of the KLK-L Related Protein; and (d) isolating the KLK-L Related Protein.

20 The invention further broadly contemplates an isolated KLK-L Protein comprising the amino acid sequence as shown in Tables 2 to 9.

The KLK-L Related Proteins of the invention may be conjugated with other molecules, such as proteins, to prepare fusion proteins. This may be accomplished, for example, by the synthesis of N-terminal or C-terminal fusion proteins.

25 The invention further contemplates antibodies having specificity against an epitope of a KLK-L Related Protein of the invention. Antibodies may be labeled with a detectable substance and used to detect proteins of the invention in tissues and cells.

The invention also permits the construction of nucleotide probes which are unique to the 30 nucleic acid molecules of the invention and/or to proteins of the invention. Therefore, the invention also relates to a probe comprising a nucleic acid sequence of the invention, or a nucleic acid sequence encoding a protein of the invention, or a part thereof. The probe may be

labeled, for example, with a detectable substance and it may be used to select from a mixture of nucleotide sequences a nucleic acid molecule of the invention including nucleic acid molecules coding for a protein which displays one or more of the properties of a protein of the invention.

5 The invention still further provides a method for identifying a substance which binds to a protein of the invention comprising reacting the protein with at least one substance which potentially can bind with the protein, under conditions which permit the formation of complexes between the substance and protein and assaying for complexes, for free substance, or for non-complexed protein. The invention also contemplates methods for identifying substances that
10 bind to other intracellular proteins that interact with a KLK-L Related Protein. Methods can also be utilized which identify compounds which bind to KLK-L gene regulatory sequences (e.g. promoter sequences).

15 Still further the invention provides a method for evaluating a compound for its ability to modulate the biological activity of a KLK-L Related Protein of the invention. For example a substance which inhibits or enhances the interaction of the protein and a substance which binds to the protein may be evaluated. In an embodiment, the method comprises providing a known concentration of a KLK-L Related Protein, with a substance which binds to the protein and a test compound under conditions which permit the formation of complexes between the substance and protein, and removing and/or detecting complexes.

20 Compounds which modulate the biological activity of a protein of the invention may also be identified using the methods of the invention by comparing the pattern and level of expression of the protein of the invention in tissues and cells, in the presence, and in the absence of the compounds.

25 The substances and compounds identified using the methods of the invention, and peptides of the invention may be used to modulate the biological activity of a KLK-L Related Protein of the invention, and they may be used in the treatment of conditions such as cancer (e.g. breast and prostate cancer). Accordingly, the substances and compounds may be formulated into compositions for administration to individuals suffering from cancer.

30 Therefore, the present invention also relates to a composition comprising one or more of a protein of the invention, a peptide of the invention, or a substance or compound identified using the methods of the invention, and a pharmaceutically acceptable carrier, excipient or

diluent. A method for treating or preventing cancer is also provided comprising administering to a patient in need thereof, a KLK-L Related Protein of the invention, or a composition of the invention.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and the specific examples while indicating preferred embodiments of the invention are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

10 BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Fig. 1. An approximate 300 Kb of contiguous genomic sequence around chromosome 19q13.3-q13.4 represented by 9 contigs, each one shown with its length in Kb. The contig numbers refer to those reported in the Lawrence Livermore National Laboratory website. Note the localization 15 of the four known genes (PSA, KLK2, Zyme, NES1). All genes are represented with arrows denoting the direction of the coding strand. The position of the stratum corneum chymotryptic enzyme (SCCE) is shown. The two genes with no homology to human kallikreins are shown as hatched arrows. The eight putative kallikrein-like genes (KLK-L1 to KLK-L8) were numbered from the most centromeric to the most telomeric. Numbers just below or just above 20 the arrows indicate Kb lengths in each contig.

DETAILED DESCRIPTION OF THE INVENTION

1. Nucleic Acid Molecules of the Invention

As hereinbefore mentioned, the invention provides an isolated nucleic acid molecule having a sequence encoding KLK-L Protein. The term "isolated" refers to a nucleic acid 25 substantially free of cellular material or culture medium when produced by recombinant DNA techniques, or chemical reactants, or other chemicals when chemically synthesized. An "isolated" nucleic acid may also be free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid molecule) from which the nucleic acid is derived. The term "nucleic acid" is intended to include DNA and RNA and can be either 30 double stranded or single stranded. In an embodiment, a nucleic acid molecule encodes a KLK-L Protein comprising the amino acid sequence as shown in Tables 2 to 9.

SEARCHED - INDEXED -

The invention includes nucleic acid sequences complementary to a nucleic acid encoding KLK-L Protein comprising the amino acid sequence as shown in Table 2 to 9.

The invention includes nucleic acid molecules having substantial sequence identity or homology to nucleic acid sequences of the invention or encoding proteins having substantial identity or similarity to the amino acid sequence shown in Tables 2 to 9. Preferably, the nucleic acids have substantial sequence identity for example at least 70% nucleic acid identity; more preferably 80% nucleic acid identity; and most preferably at least 89% to 95% sequence identity.

“Identity” as known in the art and used herein, is a relationship between two or more amino acid sequences or two or more nucleic acid sequences, as determined by comparing the sequences. It also refers to the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. Identity and similarity are well known terms to skilled artisans and they can be calculated by conventional methods (for example see Computational Molecular Biology, Lesk, A.M. ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W. ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M. and Griffin, H.G. eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G. Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J. eds. M. Stockton Press, New York, 1991, Carillo, H. and Lipman, D., SIAM J. Applied Math. 48:1073, 1988). Methods which are designed to give the largest match between the sequences are generally preferred. Methods to determine identity and similarity are codified in publicly available computer programs including the GCG program package (Devereux J. et al., Nucleic Acids Research 12(1): 387, 1984); BLASTP, BLASTN, and FASTA (Atschul, S.F. et al. J. Molec. Biol. 215: 403-410, 1990). The BLAST X program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S. et al. NCBI NLM NIH Bethesda, Md. 20894; Altschul, S. et al. J. Mol. Biol. 215: 403-410, 1990).

Isolated nucleic acid molecules encoding a protein having the activity of KLK-L Protein, and having a sequence which differs from the nucleic acid sequences of the invention due to degeneracy in the genetic code are also within the scope of the invention. Such nucleic acids encode functionally equivalent proteins (e.g., a KLK-L Protein) but differ in sequence from the sequence of a KLK-L Protein due to degeneracy in the genetic code. As one example, DNA

sequence polymorphisms within the nucleotide sequence of a KLK-L Protein may result in silent mutations which do not affect the amino acid sequence. Variations in one or more nucleotides may exist among individuals within a population due to natural allelic variation. Any and all such nucleic acid variations are within the scope of the invention. DNA sequence 5 polymorphisms may also occur which lead to changes in the amino acid sequence of a KLK-L Protein. These amino acid polymorphisms are also within the scope of the present invention.

Another aspect of the invention provides a nucleic acid molecule which hybridizes under stringent conditions, preferably high stringency conditions to a nucleic acid molecule which comprises a sequence which encodes a KLK-L Protein having the amino acid sequence shown 10 in Tables 2 to 9. Appropriate stringency conditions which promote DNA hybridization are known to those skilled in the art, or can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C may be employed. The stringency may be selected based on the conditions used in the wash step. By way of example, 15 the salt concentration in the wash step can be selected from a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be at high stringency conditions, at about 65°C.

It will be appreciated that the invention includes nucleic acid molecules encoding a KLK-L Related Protein including truncations of a KLK-L Protein, and analogs of a KLK-L Protein as described herein. It will further be appreciated that variant forms of the nucleic acid 20 molecules of the invention which arise by alternative splicing of an mRNA corresponding to a cDNA of the invention are encompassed by the invention.

An isolated nucleic acid molecule of the invention which comprises DNA can be isolated by preparing a labelled nucleic acid probe based on all or part of a nucleic acid 25 sequence of the invention. The labeled nucleic acid probe is used to screen an appropriate DNA library (e.g. a cDNA or genomic DNA library). For example, a cDNA library can be used to isolate a cDNA encoding a KLK-L Related Protein by screening the library with the labeled probe using standard techniques. Alternatively, a genomic DNA library can be similarly screened to isolate a genomic clone encompassing a gene encoding a KLK-L Related Protein. 30 Nucleic acids isolated by screening of a cDNA or genomic DNA library can be sequenced by standard techniques.

An isolated nucleic acid molecule of the invention which is DNA can also be isolated by selectively amplifying a nucleic acid encoding a KLK-L Related Protein using the polymerase chain reaction (PCR) methods and cDNA or genomic DNA. It is possible to design synthetic oligonucleotide primers from the nucleotide sequence of the invention for use in PCR.

5 A nucleic acid can be amplified from cDNA or genomic DNA using these oligonucleotide primers and standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. cDNA may be prepared from mRNA, by isolating total cellular mRNA by a variety of techniques, for example, by using the guanidinium-thiocyanate extraction procedure of Chirgwin et al., *Biochemistry*, 18, 10 5294-5299 (1979). cDNA is then synthesized from the mRNA using reverse transcriptase (for example, Moloney MLV reverse transcriptase available from Gibco/BRL, Bethesda, MD, or AMV reverse transcriptase available from Seikagaku America, Inc., St. Petersburg, FL).

An isolated nucleic acid molecule of the invention which is RNA can be isolated by cloning a cDNA encoding a KLK-L Related Protein into an appropriate vector which allows 15 for transcription of the cDNA to produce an RNA molecule which encodes a KLK-L Related Protein. For example, a cDNA can be cloned downstream of a bacteriophage promoter, (e.g. a T7 promoter) in a vector, cDNA can be transcribed *in vitro* with T7 polymerase, and the resultant RNA can be isolated by conventional techniques.

Nucleic acid molecules of the invention may be chemically synthesized using standard 20 techniques. Methods of chemically synthesizing polydeoxynucleotides are known, including but not limited to solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071).

25 Determination of whether a particular nucleic acid molecule encodes a KLK-L Related Protein can be accomplished by expressing the cDNA in an appropriate host cell by standard techniques, and testing the expressed protein in the methods described herein. A cDNA encoding a KLK-L Related Protein can be sequenced by standard techniques, such as dideoxynucleotide chain termination or Maxam-Gilbert chemical sequencing, to determine the 30 nucleic acid sequence and the predicted amino acid sequence of the encoded protein.

The initiation codon and untranslated sequences of a KLK-L Related Protein may be

determined using computer software designed for the purpose, such as PC/Gene (IntelliGenetics Inc., Calif.). The intron-exon structure and the transcription regulatory sequences of a gene encoding a *KLK-L* Related Protein may be identified by using a nucleic acid molecule of the invention encoding a *KLK-L* Related Protein to probe a genomic DNA clone library.

5 Regulatory elements can be identified using standard techniques. The function of the elements can be confirmed by using these elements to express a reporter gene such as the *lacZ* gene which is operatively linked to the elements. These constructs may be introduced into cultured cells using conventional procedures or into non-human transgenic animal models. In addition to identifying regulatory elements in DNA, such constructs may also be used to identify nuclear

10 proteins interacting with the elements, using techniques known in the art.

In a particular embodiment of the invention, the nucleic acid molecules isolated using the methods described herein are mutant *KLK-L* gene alleles. The mutant alleles may be isolated from individuals either known or proposed to have a genotype which contributes to the symptoms of cancer (e.g. breast or prostate cancer). Mutant alleles and mutant allele products

15 may be used in therapeutic and diagnostic methods described herein. For example, a cDNA of a mutant *KLK-L* gene may be isolated using PCR as described herein, and the DNA sequence of the mutant allele may be compared to the normal allele to ascertain the mutation(s) responsible for the loss or alteration of function of the mutant gene product. A genomic library can also be constructed using DNA from an individual suspected of or known to carry a mutant

20 allele, or a cDNA library can be constructed using RNA from tissue known, or suspected to express the mutant allele. A nucleic acid encoding a normal *KLK-L* gene or any suitable fragment thereof, may then be labeled and used as a probe to identify the corresponding mutant allele in such libraries. Clones containing mutant sequences can be purified and subjected to sequence analysis. In addition, an expression library can be constructed using cDNA from RNA

25 isolated from a tissue of an individual known or suspected to express a mutant *KLK-L* allele. Gene products made by the putatively mutant tissue may be expressed and screened, for example using antibodies specific for a *KLK-L* Related Protein as described herein. Library clones identified using the antibodies can be purified and subjected to sequence analysis.

The sequence of a nucleic acid molecule of the invention, or a fragment of the molecule,

30 may be inverted relative to its normal presentation for transcription to produce an antisense nucleic acid molecule. An antisense nucleic acid molecule may be constructed using chemical

SEARCHED - INDEXED
SERIALIZED - FILED

synthesis and enzymatic ligation reactions using procedures known in the art.

2. **Proteins of the Invention**

An amino acid sequence of KLK-L Protein comprises a sequence as shown in Tables 2 to 9.

5 In addition to proteins comprising an amino acid sequence as shown Tables 2 to 9 the proteins of the present invention include truncations of a KLK-L Protein, analogs of a KLK-L Protein, and proteins having sequence identity or similarity to a KLK-L Protein, and truncations thereof as described herein (i.e. KLK-L Related Proteins). Truncated proteins may comprise peptides of between 3 and 70 amino acid residues, ranging in size from a tripeptide
10 to a 70 mer polypeptide.

The truncated proteins may have an amino group (-NH2), a hydrophobic group (for example, carbobenzoyl, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end. The
15 truncated proteins may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end.

The proteins of the invention may also include analogs of a KLK-L Protein, and/or truncations thereof as described herein, which may include, but are not limited to a KLK-L
20 Protein, containing one or more amino acid substitutions, insertions, and/or deletions. Amino acid substitutions may be of a conserved or non-conserved nature. Conserved amino acid substitutions involve replacing one or more amino acids of a KLK-L Protein amino acid sequence with amino acids of similar charge, size, and/or hydrophobicity characteristics. When only conserved substitutions are made the resulting analog is preferably functionally equivalent
25 to a KLK-L Protein. Non-conserved substitutions involve replacing one or more amino acids of the KLK-L Protein amino acid sequence with one or more amino acids which possess dissimilar charge, size, and/or hydrophobicity characteristics.

One or more amino acid insertions may be introduced into a KLK-L Protein. Amino acid insertions may consist of single amino acid residues or sequential amino acids ranging from 2
30 to 15 amino acids in length.

Deletions may consist of the removal of one or more amino acids, or discrete portions

from a **KLK-L Protein** sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 10 amino acids, preferably 100 amino acids.

The proteins of the invention include proteins with sequence identity or similarity to a **KLK-L Protein** and/or truncations thereof as described herein. Such **KLK-L Proteins** include 5 proteins whose amino acid sequences are comprised of the amino acid sequences of **KLK-L Protein** regions from other species that hybridize under selected hybridization conditions (see discussion of stringent hybridization conditions herein) with a probe used to obtain a **KLK-L Protein**. These proteins will generally have the same regions which are characteristic of a **KLK-L Protein**. Preferably a protein will have substantial sequence identity for example, about 50% 10 identity, preferably 70 to 80% identity, more preferably at least 90% to 95% identity, and most preferably 98% identity with the amino acid sequence shown in Tables 2 to 9.

A percent amino acid sequence homology, similarity or identity is calculated as the 15 percentage of aligned amino acids that match the reference sequence using known methods as described herein.

The invention also contemplates isoforms of the proteins of the invention. An isoform contains the same number and kinds of amino acids as the protein of the invention, but the isoform has a different molecular structure. The isoforms contemplated by the present invention preferably have the same properties as the protein of the invention as described herein.

20 The present invention also includes **KLK-L Related Proteins** conjugated with a selected protein, or a marker protein (see below) to produce fusion proteins. Additionally, immunogenic portions of a **KLK-L Protein** and a **KLK-L Protein Related Protein** are within the scope of the invention.

A **KLK-L Related Protein** of the invention may be prepared using recombinant DNA 25 methods. Accordingly, the nucleic acid molecules of the present invention having a sequence which encodes a **KLK-L Related Protein** of the invention may be incorporated in a known manner into an appropriate expression vector which ensures good expression of the protein. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses (e.g. replication defective retroviruses, adenoviruses and adeno-associated viruses), so 30 long as the vector is compatible with the host cell used.

The invention therefore contemplates a recombinant expression vector of the invention

CONFIDENTIAL

containing a nucleic acid molecule of the invention, and the necessary regulatory sequences for the transcription and translation of the inserted protein-sequence. Suitable regulatory sequences may be derived from a variety of sources, including bacterial, fungal, viral, mammalian, or insect genes (For example, see the regulatory sequences described in Goeddel, Gene Expression 5 Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Selection of appropriate regulatory sequences is dependent on the host cell chosen as discussed below, and may be readily accomplished by one of ordinary skill in the art. The necessary regulatory sequences may be supplied by the native KLK-L Protein and/or its flanking regions.

The invention further provides a recombinant expression vector comprising a DNA 10 nucleic acid molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is linked to a regulatory sequence in a manner which allows for expression, by transcription of the DNA molecule, of an RNA molecule which is antisense to the nucleic acid sequence of a protein of the invention or a fragment thereof. Regulatory sequences linked to the antisense nucleic acid can be chosen which direct the 15 continuous expression of the antisense RNA molecule in a variety of cell types, for instance a viral promoter and/or enhancer, or regulatory sequences can be chosen which direct tissue or cell type specific expression of antisense RNA.

The recombinant expression vectors of the invention may also contain a marker gene 20 which facilitates the selection of host cells transformed or transfected with a recombinant molecule of the invention. Examples of marker genes are genes encoding a protein such as G418 and hygromycin which confer resistance to certain drugs, β -galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an immunoglobulin or portion thereof such as the Fc portion of an immunoglobulin preferably IgG. The markers can be introduced on a separate vector from the nucleic acid of interest.

25 The recombinant expression vectors may also contain genes which encode a fusion moiety which provides increased expression of the recombinant protein; increased solubility of the recombinant protein; and aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be added to the target recombinant protein to allow separation of the recombinant protein from the fusion 30 moiety subsequent to purification of the fusion protein. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly,

SEARCHED: 09/26/90
INDEXED: 09/26/90
SERIALIZED: 09/26/90
FILED: 09/26/90

MA) and pRITS (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the recombinant protein.

The recombinant expression vectors may be introduced into host cells to produce a transformant host cell. "Transformant host cells" include host cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed with", "transfected with", "transformation" and "transfection" encompass the introduction of a nucleic acid (e.g. a vector) into a cell by one of many standard techniques. Prokaryotic cells can be transformed with a nucleic acid by, for example, electroporation or calcium-chloride mediated transformation. Nucleic acid can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the proteins of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1991).

A host cell may also be chosen which modulates the expression of an inserted nucleic acid sequence, or modifies (e.g. glycosylation or phosphorylation) and processes (e.g. cleaves) the protein in a desired fashion. Host systems or cell lines may be selected which have specific and characteristic mechanisms for post-translational processing and modification of proteins. For example, eukaryotic host cells including CHO, VERO, BHK, HeLa, COS, MDCK, 293, 25 3T3, and WI38 may be used. For long-term high-yield stable expression of the protein, cell lines and host systems which stably express the gene product may be engineered.

Host cells and in particular cell lines produced using the methods described herein may be particularly useful in screening and evaluating compounds that modulate the activity of a KLK-L Related Protein.

30 The proteins of the invention may also be expressed in non-human transgenic animals including but not limited to mice, rats, rabbits, guinea pigs, micro-pigs, goats, sheep, pigs, non-

GOVERNMENT EDITION

human primates (e.g. baboons, monkeys, and chimpanzees) [see Hammer et al. (Nature 315:680-683, 1985), Palmiter et al. (Science 222:809-814, 1983), Brinster et al. (Proc Natl. Acad. Sci USA 82:44384442, 1985), Palmiter and Brinster (Cell. 41:343-345, 1985) and U.S. Patent No. 4,736,866)]. Procedures known in the art may be used to introduce a nucleic acid 5 molecule of the invention encoding a *KLK-L* Related Protein into animals to produce the founder lines of transgenic animals. Such procedures include pronuclear microinjection, retrovirus mediated gene transfer into germ lines, gene targeting in embryonic stem cells, electroporation of embryos, and sperm-mediated gene transfer.

The present invention contemplates a transgenic animal that carries the *KLK-L* gene in 10 all their cells, and animals which carry the transgene in some but not all their cells. The transgene may be integrated as a single transgene or in concatamers. The transgene may be selectively introduced into and activated in specific cell types (See for example, Lasko et al, 1992 Proc. Natl. Acad. Sci. USA 89: 6236). The transgene may be integrated into the chromosomal site of the endogenous gene by gene targeting. The transgene may be selectively 15 introduced into a particular cell type inactivating the endogenous gene in that cell type (See Gu et al Science 265: 103-106).

The expression of a recombinant *KLK-L* Related Protein in a transgenic animal may be assayed using standard techniques. Initial screening may be conducted by Southern Blot 20 analysis, or PCR methods to analyze whether the transgene has been integrated. The level of mRNA expression in the tissues of transgenic animals may also be assessed using techniques including Northern blot analysis of tissue samples, *in situ* hybridization, and RT-PCR. Tissue may also be evaluated immunocytochemically using antibodies against *KLK-L* Protein.

Proteins of the invention may also be prepared by chemical synthesis using techniques 25 well known in the chemistry of proteins such as solid phase synthesis (Merrifield, 1964, J. Am. Chem. Assoc. 85:2149-2154) or synthesis in homogenous solution (Houbenweyl, 1987, Methods of Organic Chemistry, ed. E. Wansch, Vol. 15 I and II, Thieme, Stuttgart).

N-terminal or C-terminal fusion proteins comprising a *KLK-L* Related Protein of the 30 invention conjugated with other molecules, such as proteins, may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of a *KLK-L* Related Protein, and the sequence of a selected protein or marker protein with a desired biological function. The resultant fusion proteins contain *KLK-L* Protein fused to the selected protein or marker protein as

described herein. Examples of proteins which may be used to prepare fusion proteins include immunoglobulins, glutathione-S-transferase (GST), hemagglutinin (HA), and truncated myc.

3. Antibodies

KLK-L Related Proteins of the invention can be used to prepare antibodies specific for the proteins. Antibodies can be prepared which bind a distinct epitope in an unconserved region of the protein. An unconserved region of the protein is one that does not have substantial sequence homology to other proteins. A region from a conserved region such as a well-characterized domain can also be used to prepare an antibody to a conserved region of a KLK-L Related Protein. Antibodies having specificity for a KLK-L Related Protein may also be raised from fusion proteins created by expressing fusion proteins in bacteria as described herein.

The invention can employ intact monoclonal or polyclonal antibodies, and immunologically active fragments (e.g. a Fab, (Fab)₂ fragment, or Fab expression library fragments and epitope-binding fragments thereof), an antibody heavy chain, and antibody light chain, a genetically engineered single chain Fv molecule (Ladner et al, U.S. Pat. No. 4,946,778), or a chimeric antibody, for example, an antibody which contains the binding specificity of a murine antibody, but in which the remaining portions are of human origin. Antibodies including monoclonal and polyclonal antibodies, fragments and chimeras, may be prepared using methods known to those skilled in the art.

4. Applications of the Nucleic Acid Molecules, KLK-L Related Proteins, and Antibodies of the Invention

The nucleic acid molecules, KLK-L Related Proteins, and antibodies of the invention may be used in the prognostic and diagnostic evaluation of cancer (e.g. breast and prostate cancer), and the identification of subjects with a predisposition to cancer (Section 4.1.1 and 4.1.2). Methods for detecting nucleic acid molecules and KLK-L Related Proteins of the invention, can be used to monitor cancer by detecting KLK-L Related Proteins and nucleic acid molecules encoding KLK-L Related Proteins. It would also be apparent to one skilled in the art that the methods described herein may be used to study the developmental expression of KLK-L Related Proteins and, accordingly, will provide further insight into the role of KLK-L Related Proteins. The applications of the present invention also include methods for the identification of compounds that modulate the biological activity of KLK-L or KLK-L Related Proteins (Section 4.2). The compounds, antibodies etc. may be used for the treatment of cancer (Section

4.3).

4.1 Diagnostic Methods

A variety of methods can be employed for the diagnostic and prognostic evaluation of cancer (e.g. breast and prostate cancer), and the identification of subjects with a predisposition to cancer. Such methods may, for example, utilize nucleic acid molecules of the invention, and fragments thereof, and antibodies directed against *KLK-L* Related Proteins, including peptide fragments. In particular, the nucleic acids and antibodies may be used, for example, for: (1) the detection of the presence of *KLK-L* mutations, or the detection of either over- or under-expression of *KLK-L* mRNA relative to a non-disorder state or the qualitative or quantitative detection of alternatively spliced forms of *KLK-L* transcripts which may correlate with certain conditions or susceptibility toward such conditions; and (2) the detection of either an over- or an under-abundance of *KLK-L* Related Proteins relative to a non- disorder state or the presence of a modified (e.g., less than full length) *KLK-L* Protein which correlates with a disorder state, or a progression toward a disorder state.

15 The methods described herein may be performed by utilizing pre-packaged diagnostic kits comprising at least one specific *KLK-L* nucleic acid or antibody described herein, which may be conveniently used, e.g., in clinical settings, to screen and diagnose patients and to screen and identify those individuals exhibiting a predisposition to developing a disorder.

Nucleic acid-based detection techniques are described, below, in Section 4.1.1. Peptide detection techniques are described, below, in Section 4.1.2. The samples that may be analyzed using the methods of the invention include those which are known or suspected to express *KLK-L* or contain *KLK-L* Related Proteins. The samples may be derived from a patient or a cell culture, and include but are not limited to biological fluids, tissue extracts, freshly harvested cells, and lysates of cells which have been incubated in cell cultures.

4.1.1 Methods for Detecting Nucleic Acid Molecules of the Invention

The nucleic acid molecules of the invention allow those skilled in the art to construct nucleotide probes for use in the detection of nucleic acid sequences of the invention in samples. Suitable probes include nucleic acid molecules based on nucleic acid sequences encoding at least 5 sequential amino acids from regions of the KLK-L Protein, preferably they comprise 15 to 30 nucleotides. A nucleotide probe may be labeled with a detectable substance such as a radioactive label which provides for an adequate signal and has sufficient half-life such as ^{32}P ,

³H, ¹⁴C or the like. Other detectable substances which may be used include antigens that are recognized by a specific labeled antibody, fluorescent compounds, enzymes, antibodies specific for a labeled antigen, and luminescent compounds. An appropriate label may be selected having regard to the rate of hybridization and binding of the probe to the nucleotide to be detected and the amount of nucleotide available for hybridization. Labeled probes may be hybridized to nucleic acids on solid supports such as nitrocellulose filters or nylon membranes as generally described in Sambrook et al, 1989, Molecular Cloning, A Laboratory Manual (2nd ed.). The nucleic acid probes may be used to detect genes, preferably in human cells, that encode KLK-L Related Proteins. The nucleotide probes may also be useful in the diagnosis of cancer; in monitoring the progression of cancer; or monitoring a therapeutic treatment.

The probe may be used in hybridization techniques to detect genes that encode KLK-L Related Proteins. The technique generally involves contacting and incubating nucleic acids (e.g. recombinant DNA molecules, cloned genes) obtained from a sample from a patient or other cellular source with a probe of the present invention under conditions favorable for the specific annealing of the probes to complementary sequences in the nucleic acids. After incubation, the non-annealed nucleic acids are removed, and the presence of nucleic acids that have hybridized to the probe if any are detected.

The detection of nucleic acid molecules of the invention may involve the amplification of specific gene sequences using an amplification method such as PCR, followed by the analysis of the amplified molecules using techniques known to those skilled in the art. Suitable primers can be routinely designed by one of skill in the art.

Genomic DNA may be used in hybridization or amplification assays of biological samples to detect abnormalities involving *klk-1* structure, including point mutations, insertions, deletions, and chromosomal rearrangements. For example, direct sequencing, single stranded conformational polymorphism analyses, heteroduplex analysis, denaturing gradient gel electrophoresis, chemical mismatch cleavage, and oligonucleotide hybridization may be utilized.

Genotyping techniques known to one skilled in the art can be used to type polymorphisms that are in close proximity to the mutations in a *klk-1* gene. The polymorphisms may be used to identify individuals in families that are likely to carry mutations. If a polymorphism exhibits linkage disequilibrium with mutations in a *klk-1* gene, it can also be used to screen for individuals in the general population likely to carry mutations. Polymorphisms

which may be used include restriction fragment length polymorphisms (RFLPs), single-base polymorphisms, and simple sequence repeat polymorphisms (SSLPs).

A probe of the invention may be used to directly identify RFLPs. A probe or primer of the invention can additionally be used to isolate genomic clones such as YACs, BACs, PACs, 5 cosmids, phage or plasmids. The DNA in the clones can be screened for SSLPs using hybridization or sequencing procedures.

Hybridization and amplification techniques described herein may be used to assay qualitative and quantitative aspects of *klk-1* expression. For example, RNA may be isolated from a cell type or tissue known to express *klk-1* and tested utilizing the hybridization (e.g. standard 10 Northern analyses) or PCR techniques referred to herein. The techniques may be used to detect differences in transcript size which may be due to normal or abnormal alternative splicing. The techniques may be used to detect quantitative differences between levels of full length and/or alternatively splice transcripts detected in normal individuals relative to those individuals exhibiting cancer symptoms or other disease conditions.

15 The primers and probes may be used in the above described methods *in situ* i.e directly on tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections.

4.1.2 Methods for Detecting KLK-L Related Proteins

Antibodies specifically reactive with a KLK-L Related Protein, or derivatives, such as 20 enzyme conjugates or labeled derivatives, may be used to detect KLK-L Related Proteins in various samples (e.g. biological materials). They may be used as diagnostic or prognostic reagents and they may be used to detect abnormalities in the level of KLK-L Related Proteins expression, or abnormalities in the structure, and/or temporal, tissue, cellular, or subcellular 25 location of a KLK-L Related Protein. Antibodies may also be used to screen potentially therapeutic compounds *in vitro* to determine their effects on cancer, and other conditions. *In vitro* immunoassays may also be used to assess or monitor the efficacy of particular therapies. The antibodies of the invention may also be used *in vitro* to determine the level of KLK-L expression in cells genetically engineered to produce a KLK-L Related Protein.

The antibodies may be used in any known immunoassays which rely on the binding 30 interaction between an antigenic determinant of a KLK-L Related Protein and the antibodies. Examples of such assays are radioimmunoassays, enzyme immunoassays (e.g. ELISA), immunofluorescence, immunoprecipitation, latex agglutination, hemagglutination, and

histochemical tests. The antibodies may be used to detect and quantify KLK-L Related Proteins in a sample in order to determine its role in particular cellular events or pathological states, and to diagnose and treat such pathological states.

5 In particular, the antibodies of the invention may be used in immuno-histochemical analyses, for example, at the cellular and sub-subcellular level, to detect a KLK-L Related Protein, to localize it to particular cells and tissues, and to specific subcellular locations, and to quantitate the level of expression.

10 Cytochemical techniques known in the art for localizing antigens using light and electron microscopy may be used to detect a KLK-L Related Protein. Generally, an antibody of the invention may be labeled with a detectable substance and a KLK-L Related Protein may be localised in tissues and cells based upon the presence of the detectable substance. Examples 15 of detectable substances include, but are not limited to, the following: radioisotopes (e.g., ^3H , ^{14}C , ^{35}S , ^{125}I , ^{131}I), fluorescent labels (e.g., FITC, rhodamine, lanthanide phosphors), luminescent labels such as luminol; enzymatic labels (e.g., horseradish peroxidase, beta-galactosidase, 20 luciferase, alkaline phosphatase, acetylcholinesterase), biotinyl groups (which can be detected by marked avidin e.g., streptavidin containing a fluorescent marker or enzymatic activity that can be detected by optical or calorimetric methods), predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, metal binding domains, epitope tags). In some embodiments, labels are attached via spacer arms of various lengths to reduce potential steric hindrance. Antibodies may also be coupled to electron dense substances, such as ferritin or colloidal gold, which are readily visualised by electron microscopy.

25 The antibody or sample may be immobilized on a carrier or solid support which is capable of immobilizing cells, antibodies etc. For example, the carrier or support may be nitrocellulose, or glass, polyacrylamides, gabbros, and magnetite. The support material may have any possible configuration including spherical (e.g. bead), cylindrical (e.g. inside surface of a test tube or well, or the external surface of a rod), or flat (e.g. sheet, test strip). Indirect methods may also be employed in which the primary antigen-antibody reaction is amplified by the introduction of a second antibody, having specificity for the antibody reactive against KLK- 30 L Related Protein. By way of example, if the antibody having specificity against a KLK-L Related Protein is a rabbit IgG antibody, the second antibody may be goat anti-rabbit gamma-

globulin labeled with a detectable substance as described herein.

Where a radioactive label is used as a detectable substance, a KLK-L Related Protein may be localized by radioautography. The results of radioautography may be quantitated by determining the density of particles in the radioautographs by various optical methods, or by counting the grains.

5 **4.2 Methods for Identifying or Evaluating Substances/Compounds**

The methods described herein are designed to identify substances that modulate the biological activity of a KLK-L Related Protein including substances that bind to KLK-L Related Proteins, or bind to other proteins that interact with a KLK-L Related Protein, to compounds that interfere with, or enhance the interaction of a KLK-L Related Protein and substances that bind to the KLK-L Related Protein or other proteins that interact with a KLK-L Related Protein. Methods are also utilized that identify compounds that bind to *KLK-L* regulatory sequences.

10 The substances and compounds identified using the methods of the invention include but are not limited to peptides such as soluble peptides including Ig-tailed fusion peptides, members of random peptide libraries and combinatorial chemistry-derived molecular libraries made of D- and/or L-configuration amino acids, phosphopeptides (including members of random or partially degenerate, directed phosphopeptide libraries), antibodies [e.g. polyclonal, 15 monoclonal, humanized, anti-idiotypic, chimeric, single-chain antibodies; fragments, (e.g. Fab, Fab₂, and Fab expression library fragments, and epitope-binding fragments thereof)], and small organic or inorganic molecules. The substance or compound may be an endogenous 20 physiological compound or it may be a natural or synthetic compound.

25 Substances which modulate a KLK-L Related Protein can be identified based on their ability to bind to a KLK-L Related Protein. Therefore, the invention also provides methods for identifying substances which bind to a KLK-L Related Protein. Substances identified using the methods of the invention may be isolated, cloned and sequenced using conventional techniques.

30 Substances which can bind with a KLK-L Related Protein may be identified by reacting a KLK-L Related Protein with a test substance which potentially binds to a KLK-L Related Protein, under conditions which permit the formation of substance-KLK-L Related Protein complexes and removing and/or detecting the complexes. The complexes can be detected by assaying for substance-KLK-L Related Protein complexes, for free substance, or for non-

complexed KLK-L Related Protein. Conditions which permit the formation of substance-KLK-L Related Protein complexes may be selected having regard to factors such as the nature and amounts of the substance and the protein.

The substance-protein complex, free substance or non-complexed proteins may be isolated by conventional isolation techniques, for example, salting out, chromatography, electrophoresis, gel filtration, fractionation, absorption, polyacrylamide gel electrophoresis, agglutination, or combinations thereof. To facilitate the assay of the components, antibody against KLK-L Related Protein or the substance, or labeled KLK-L Related Protein, or a labeled substance may be utilized. The antibodies, proteins, or substances may be labeled with a detectable substance as described above.

A KLK-L Related Protein, or the substance used in the method of the invention may be insolubilized. For example, a KLK-L Related Protein, or substance may be bound to a suitable carrier such as agarose, cellulose, dextran, Sephadex, Sepharose, carboxymethyl cellulose polystyrene, filter paper, ion-exchange resin, plastic film, plastic tube, glass beads, polyamine-methyl vinyl-ether-maleic acid copolymer, amino acid copolymer, ethylene-maleic acid copolymer, nylon, silk, etc. The carrier may be in the shape of, for example, a tube, test plate, beads, disc, sphere etc. The insolubilized protein or substance may be prepared by reacting the material with a suitable insoluble carrier using known chemical or physical methods, for example, cyanogen bromide coupling.

The invention also contemplates a method for evaluating a compound for its ability to modulate the biological activity of a KLK-L Related Protein of the invention, by assaying for an agonist or antagonist (i.e. enhancer or inhibitor) of the binding of a KLK-L Related Protein with a substance which binds with a KLK-L Related Protein. The basic method for evaluating if a compound is an agonist or antagonist of the binding of a KLK-L Related Protein and a substance that binds to the protein, is to prepare a reaction mixture containing the KLK-L Related Protein and the substance under conditions which permit the formation of substance-KLK-L Related Protein complexes, in the presence of a test compound. The test compound may be initially added to the mixture, or may be added subsequent to the addition of the KLK-L Related Protein and substance. Control reaction mixtures without the test compound or with a placebo are also prepared. The formation of complexes is detected and the formation of complexes in the control reaction but not in the reaction mixture indicates that the test

SEARCHED
INDEXED
SERIALIZED
FILED

compound interferes with the interaction of the KLK-L Related Protein and substance. The reactions may be carried out in the liquid phase or the KLK-L Related Protein, substance, or test compound may be immobilized as described herein. The ability of a compound to modulate the biological activity of a KLK-L Related Protein of the invention may be tested by 5 determining the biological effects on cells.

It will be understood that the agonists and antagonists i.e. inhibitors and enhancers that can be assayed using the methods of the invention may act on one or more of the binding sites on the protein or substance including agonist binding sites, competitive antagonist binding sites, non-competitive antagonist binding sites or allosteric sites.

10 The invention also makes it possible to screen for antagonists that inhibit the effects of an agonist of the interaction of KLK-L Related Protein with a substance which is capable of binding to the KLK-L Related Protein. Thus, the invention may be used to assay for a compound that competes for the same binding site of a KLK-L Related Protein.

15 The invention also contemplates methods for identifying compounds that bind to proteins that interact with a KLK-L Related Protein. Protein-protein interactions may be identified using conventional methods such as co-immunoprecipitation, crosslinking and co-purification through gradients or chromatographic columns. Methods may also be employed that result in the simultaneous identification of genes which encode proteins interacting with a KLK-L Related Protein. These methods include probing expression libraries with labeled KLK-L 20 Related Protein.

Two-hybrid systems may also be used to detect protein interactions *in vivo*. Generally, 25 plasmids are constructed that encode two hybrid proteins. A first hybrid protein consists of the DNA-binding domain of a transcription activator protein fused to a KLK-L Related Protein, and the second hybrid protein consists of the transcription activator protein's activator domain fused to an unknown protein encoded by a cDNA which has been recombined into the plasmid as part of a cDNA library. The plasmids are transformed into a strain of yeast (e.g. *S. cerevisiae*) that contains a reporter gene (e.g. lacZ, luciferase, alkaline phosphatase, horseradish peroxidase) 30 whose regulatory region contains the transcription activator's binding site. The hybrid proteins alone cannot activate the transcription of the reporter gene. However, interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product.

It will be appreciated that fusion proteins may be used in the above-described methods. In particular, KLK-L Related Proteins fused to a glutathione-S-transferase may be used in the methods.

The reagents suitable for applying the methods of the invention to evaluate compounds 5 that modulate a KLK-L Related Protein may be packaged into convenient kits providing the necessary materials packaged into suitable containers. The kits may also include suitable supports useful in performing the methods of the invention.

4.3 Compositions and Treatments

The substances or compounds identified by the methods described herein, antibodies, 10 and antisense nucleic acid molecules of the invention, and peptides may be used for modulating the biological activity of a KLK-L Related Protein, and they may be used in the treatment of conditions such as cancer (e.g. prostate or breast cancer). Accordingly, the substances, antibodies, peptides, and compounds may be formulated into pharmaceutical compositions for administration to subjects in a biologically compatible form suitable for administration *in vivo*. 15 By "biologically compatible form suitable for administration *in vivo*" is meant a form of the active substance to be administered in which any toxic effects are outweighed by the therapeutic effects. The active substances may be administered to living organisms including humans, and animals. Administration of a therapeutically active amount of a pharmaceutical composition of the present invention is defined as an amount effective, at dosages and for periods of time 20 necessary to achieve the desired result. For example, a therapeutically active amount of a substance may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of antibody to elicit a desired response in the individual. Dosage regima may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated 25 by the exigencies of the therapeutic situation.

The active substance may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or rectal administration. Depending on the route of administration, the active substance may be coated in a material to protect the substance from the action of enzymes, acids and other natural 30 conditions that may inactivate the substance.

The compositions described herein can be prepared by *per se* known methods for the

preparation of pharmaceutically acceptable compositions which can be administered to subjects, such that an effective quantity of the active substance is combined in a mixture with a pharmaceutically acceptable vehicle. Suitable vehicles are described, for example, in Remington's Pharmaceutical Sciences (Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, Pa., USA 1985). On this basis, the compositions include, albeit not exclusively, solutions of the active substances in association with one or more pharmaceutically acceptable vehicles or diluents, and contained in buffered solutions with a suitable pH and isosmotic with the physiological fluids.

The activity of the substances, compounds, antibodies, antisense nucleic acid molecules, and compositions of the invention may be confirmed in animal experimental model systems.

The following non-limiting example is illustrative of the present invention:

Example

MATERIALS AND METHODS

Identification of positive PAC and BAC genomic clones from a human genomic DNA

15 **library.**

The sequence of PSA, KLK1, KLK2, NES1 and Zyme genes is already known. Polymerase chain reaction (PCR)-based amplification protocols have been developed which allowed generation of PCR products specific for each one of these genes. Using these PCR products as probes, labeled with ^{32}P , a human genomic DNA PAC library and a human genomic DNA BAC library was screened for the purpose of identifying positive clones of approximately 100-150 Kb long. The general strategies for these experiments have been published elsewhere (14). The genomic libraries were spotted in duplicate on nylon membranes and positive clones were further confirmed by Southern blot analysis as described (14).

DNA sequences on chromosome 19

25 The Lawrence Livermore National Laboratory participates in the sequencing of the human genome project and focuses on sequencing chromosome 19. Large sequencing information on this chromosome is available at the website of the Lawrence Livermore National Laboratory (<http://www-bio.llnl.gov/genome/gemnoma.html>).

30 Approximately 300 Kb of genomic sequences was obtained from that website, encompassing a region on chromosome 19q13.3 - 13.4, where the known kallikrein genes are localized. This 300 Kb of sequence is represented by 9 contigs of variable lengths. By using

a number of different computer programs, an almost contiguous sequence of the region was established as shown diagrammatically in **Figure 1**. Some of the contigs were reversed as shown in **Figure 1** in order to reconstruct the area on both strands of DNA.

By using the published sequences of PSA, KLK2, NES1 and protease M and the computer software BLAST 2, using alignment strategies, the relative positions of these genes on the contiguous map were identified (**Figure 1**). These known genes served as hallmarks for further studies. An EcoR1 restriction map of the area is also available at the website of the Lawrence Livermore National Laboratory. Using this restriction map and the computer program WebCutter (<http://www.firstmarket.com/cutter/cut2.html>), a restriction study analysis of the available sequence was performed to further confirm the assignment and relative positions of these contigs along chromosome 19. The obtained configuration is presented in **Figure 1**.

Gene prediction analysis

For exon prediction analysis of the whole genomic area, a number of different computer programs were used. These programs are listed in **Table 1**. All these programs were initially tested using known genomic sequences of the PSA, protease M and NES1 genes. The more reliable computer programs, GeneBuilder (gene prediction), GeneBuilder (exon prediction), Grail 2 and GENEID-3 were selected for further use.

Protein homology searching

Putative exons of the new genes were first translated to the corresponding aminoacid sequences. BLAST homology searching for the proteins encoded by the exons of the putative new genes were performed using the BLASTP program and the Genbank databases.

RESULTS

Relative position of PSA, KLK2, Zyme and NES1 on Chromosome 19

Screening of the human BAC library identified two clones which were positive for the Zyme gene (clones BAC 288H1 and BAC 76F7). These BACs were further analyzed by PCR and primers specific for PSA, NES1, KLK1 and KLK2. These analyses indicated that both BACs were positive for Zyme, PSA and KLK2 and negative for KLK1 and NES1 genes.

Screening of the human PAC genomic library identified a PAC clone which was positive for NES1 (clone PAC 34B1). Further PCR analysis indicated that this PAC clone was positive for NES1 and KLK1 genes and negative for PSA, KLK2 and Zyme. Combination of this information with the EcoR1 restriction map of the region allowed establishment of the

© 2002 Blackwell Science Ltd, *Journal of Internal Medicine* 252: 24–32

relative positions of these four genes. PSA is the most centromeric, followed by KLK2, Zyme and NES1. Further alignment of the known sequences of these genes with the 300 Kb contig enabled precise localization of all four genes and determination of the direction of transcription, as shown by the arrows in **Figure 1**. The KLK1 gene sequence was not identified on this contig 5 and appears to be further telomeric to NES1.

Identification of new genes

A set of rules was used to consider presence of a new gene in the genomic area of interest as follows:

1. Clusters of at least 3 exons should be found.
- 10 2. Only exons with high prediction score ("good" or "excellent" quality, as indicated by the searching programs) were considered for the construction of the putative new genes.
3. Exons predicted were reliable only if they were identified by at least two different exon prediction programs.

By using this strategy, eleven putative new genes were identified of which one was found 15 on subsequent homology analysis to be a known gene, the human stratum corneum chymotrypsin enzyme (HSCCE). Two other putative genes (gene UG-1 and gene UG-2) were identified which show no homology, at the protein level, with the kallikrein proteins. The eight remaining genes all have variable homologies with known human or animal-kallikrein proteins and/or other known serine proteases.

20 In **Tables 2 to 11**, the preliminary exon structure and encoded protein for each one of the ten newly identified genes is shown. In **Table 12**, some proteins are presented which appear, on preliminary analysis, to be homologous to the proteins encoded by the putative new genes.

DISCUSSION

25 Prediction of protein-coding genes in newly sequenced DNA becomes very important after the establishment of large genome sequencing projects. This problem is complicated due to the exon-intron structure of the eukaryotic genes which interrupts the coding sequence in many unequal parts. In order to predict the protein-coding exons and overall gene structure, a number of computer programs were developed. All these programs are based on the 30 combination of potential functional signals with the global statistical properties of known protein-coding regions (15). However, the most powerful approach for gene structure prediction

is to combine information about potential functional signals (splice sites, translation start or stop signal etc.) together with the statistical properties of coding sequences (coding potential) along with information about homologies between the predicted protein and known protein families (16).

5 In mouse and rat, kallikreins are encoded by large multigene families and these genes tend to cluster in groups with a distance as small as 3.3 – 7.0 Kb (3). A strong conservation of gene order between human chromosome 19q13.1 – q13.4 and 17 loci in a 20-cM proximal part of mouse chromosome 7, including the kallikrein locus, has been documented (17).

10 In humans, only a few kallikrein genes were identified. In fact, only KLK1, KLK2 and KLK3 (PSA) are considered to represent the human kallikrein gene family (9). In this paper, we provide strong evidence that a large number of kallikrein-like genes are clustered within a 300Kb region around chromosome 19q13.2 – q13.4. Except of the three established human kallikreins (KLK1, KLK2, KLK3), Zyme and NES1, as well as the stratum corneum chymotryptic enzyme and another eight new genes , KLK-L1 to KLK-L8, may constitute a 15 large gene family . This will bring the total number of kallikrein or kallikrein-like genes in humans to fourteen.

20 Kallikrein genes are a subfamily of serine proteases, traditionally characterized by their ability to liberate lysyl-bradykinin (kallidin) from kininogen (18). More recently, however, a new, structural concept has emerged to describe kallikreins. From accumulated sequence data, it is now clear that the mouse has many genes with high homology to kallikrein coding sequences (19-20). Richard and co-workers have contributed to the concept of a " kallikrein multigene family" to refer to these genes (21-22). This definition is not based much on specific enzymatic function of the gene product, but more on its sequence homology and their close linkage on mouse chromosome 7. In humans, only KLK1 meets the functional definition of a 25 kallikrein. KLK2 has trypsin-like enzymatic activity and KLK3 (PSA) has very weak chymotrypsin-like enzymatic activity. These activities of KLK2 and KLK3 are not known to liberate biologically active peptides from precursors. Based on the newer definition, members of the kallikrein family include, not only the gene for the kallikrein enzyme, but also genes encoding other homologous proteases, including the enzyme that processes the precursors of the 30 nerve growth factor and epidermal growth factor (8). Therefore, it is important to note the clear distinction between the enzyme kallikrein and a kallikrein or a kallikrein-like gene .

In order to test the accuracy of the computer programs, known genomic areas containing the PSA, Zyme and KLK2 genes were tested. Two of these programs (Grail 2 and GeneBuilder) were able to detect about 95% of the tested known genes (data not shown). Matches with expressed sequence tag sequences (EST) can also be employed for gene structure prediction in 5 the GeneBuilder program and this can significantly improve the power of the program especially at high stringency (e.g. >95% homology).

In mouse, ten of the kallikrein genes appear to be pseudogenes (9). Two of the new genes do not show homology with the kallikrein genes (UG-1 and UG-2). However, one of them (UG-2) is related to mouse myelin associated glycoprotein (Table 12). There may still be 10 an association between UG-2 and the kallikrein genes since some mouse kallikreins are related to nerve growth factor, as discussed earlier (8) and Zyme was found to be highly expressed in brain tissue and is claimed to be related to Alzheimer's disease (11).

Having illustrated and described the principles of the invention in a preferred 15 embodiment, it should be appreciated by those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. All modifications coming within the scope of the following claims are claimed.

All publications, patents and patent applications referred to herein are incorporated by 20 reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

FULL CITATIONS FOR REFERENCES REFERRED TO IN THE SPECIFICATION

REFERENCES

1. Evans BAE, Yun ZX, Close JA, Tregebar GW, Kitamura N, Nakanishi S, et al. Structure and chromosomal localization of the human renal kallikrein gene. *Biochemistry* 1988;27:3124-3129.
- 5 2. Clements JA. The glandular kallikrein family of enzymes: Tissue-specific and hormonal regulation. *Endocr Rev* 1989;10:393-419.
3. Evans BA, Drinkwater CC, Richards RI. Mouse glandular kallikrein genes: Structure and partial sequence analysis of the kallikrein gene locus. *J Biol Chem* 1987;262:8027-10 8034.
4. Drinkwater CC, Evans BA, Richards RI. Kallikreins, kinins and growth factor biosynthesis. *Trends Biochem Sci* 1988b;13:169-172.
5. Ashley PL, MacDonald RJ. Tissue-specific expression of kallikrein-related genes in the 15 rat. *Biochemistry* 1985;24:4520-5427.
6. Gerald WL, Chao J, Chao L. Sex dimorphism and hormonal regulation of rat tissue kallikrein mRNA. *Biochim Biophys Acta* 1986;867:16-23.
7. Riegman PHJ, Vlietstra RJ, van der Korput JAGM, Romijn JC, Trapman J. Characterization of the prostate-specific antigen gene: A novel human kallikrein-like 20 gene. *Biochem Biophys Res Commun* 1989;159:95-102.
8. Schedlich LJ, Bennetts BH, Morris BJ. Primary structure of a human glandular kallikrein gene. *DNA* 1987;6:429-437.
9. Riegman PH, Vlietstra RJ, Suurmeijer L, Cleutjens CBJM, Trapman J. Characterization 25 of the human kallikrein locus. *Genomics* 1992;14:6-11.
10. Anisowicz A, Sotiropoulou G, Stenman G, Mok SC, Sager R. A novel protease homolog differentially expressed in breast and ovarian cancer. *Mol Med* 1996;2:624-636.
11. Little SP, Dixon EP, Norris F, Buckley W, Becker GW, Johnson M, et al. Zyme, a novel 30 and potentially amyloidogenic enzyme cDNA isolated from Alzheimer's disease brain. *J Biol Chem* 1997;272:25135-25142.
12. Yamashiro K, Tsuruoka N, Kodama S, Tsujimoto M, Yamamura Y, Tanaka T, et al.

Molecular cloning of a novel trypsin-like serine protease (neurosin) preferentially expressed in brain. *Biochim Biophys Acta* 1997;1350:11-14.

13. Liu XL, Wazer DE, Watanabe K, Band V. Identification of a novel serine protease-like gene, the expression of which is down-regulated during breast cancer progression. *Cancer Res* 1996;56:3371-3379.

5 14. Luo L, Herbrick J-A, Scherer SW, Beatty B, Squire J, Diamandis EP. Structural characterization and mapping of the normal epithelial cell-specific 1 gene. *Biochem Biophys Res Commun* 1998;247:580-586.

10 15. Milanesi L, Kolchanov N, Rogozin I, Kel A, Titov I. Sequence functional inference. In: "Guide to human genome computing", ed. M.J. Bishop, Academic Press, Cambridge, 1994, 249-312.

16. Burset M, Guigo R. Evaluation of gene structure prediction programs. *Genomics* 1996;34:353-367.

15 17. Nadeau J, Grant P, Kosowsky M. Mouse and human homology map. *Mouse Genome* 1991;89:31-36.

18. Schachter M. Kallikreins (kininogenases) - a group of serine proteases with bioregulatory actions. *Pharmacol Rev* 1980;31:1-17.

19. Morris BJ, Catanzaro DF, Richards RI, Mason AJ, Shine J. Kallikrein and renin: Molecular biology and biosynthesis. *Clin Sci* 1981;61:351s-353s.

20 20. Richards RI, Catanzaro DF, Mason AJ, Morris BJ, Baxter JD, Shine J. Mouse glandular kallikrein genes. Nucleotide sequence of cloned cDNA coding for a member of the kallikrein arginyl estero-peptidase group of serine proteases. *J Biol Chem* 1982;257:2758-2761.

21. Van Leeuwen BH, Evans BA, Tregebar GW, Richards RI. Mouse glandular kallikrein genes. Identification, structure and expression of the renal kallikrein gene. *J Biol Chem* 1986;261:5529-5535.

25 22. Evans BA, Richards RI. Genes for the α and γ subunits of mouse nerve growth factor. *EMBO J* 1985;4:133-138.

23. Rogozin IB, Milanesi L, Kolchanov NA. Gene structure prediction using information on homologous protein sequence. *Comput Applic Biosci* 1996;12:161-170.

30 24. Riegman PHJ, Vlietstra RJ, Klaassen P, van de Korput JAGM, Geurts van Kessel A,

Romijn JC, et al. The prostate-specific antigen gene and the human glandular kallikrein-1 gene are tandemly located on chromosome 19. 1989;247:123-126.

Table 1. Exon or gene prediction programs used in this study.¹

No.	Program name	Source	Website or e-mail address
1	GeneBuilder (gene prediction)	Institute of Advanced Biomedical Technologies	http://125.itba.mi.cnr.it/~webgene/genebuilder.html
2	GeneBuilder(exon prediction)	Institute of Advanced Biomedical Technologies	http://125.itba.mi.cnr.it/~webgene/genebuilder.html
3	ORF gene	Institute of Advanced Biomedical Technologies	http://125.itba.mi.cnr.it/~webgene/wwworfgene2.html
4	GENEID-3	BioMolecular Engineering Research Center, Boston University	http://apolo.imim.es/geneid.html (geneid@darwin.bu.edu)
5	Grail 2	Oak Ridge National Laboratory	http://compbio.ornl.gov
6	FGENEH	Baylor College of Medicine, Houston, Texas	http://mcrb.bcm.tmc.edu

1. In the final analysis of the sequences we used programs 1, 2, 4 and 5 only.

Table 2. Predicted exons of the putative gene KLK-L1. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹		No. of bases	Translated protein sequence	Exon prediction program ²
	From(bp)	To (bp)			
1	2263	2425	162	SLVSGSCSQIINGEDCSPHSQPWQAALVMENELFCSGVLVHPQWVLSAAHCFQ	A,B,D
2	2847	3109	262	SYTIGLGLHSLEADQEPEGSQMVEASLSVRHPEYNRPLL ANDLMLIKLDESVSESDTIRSIASQCPTAGNSCLVSG WGLLANGEIT	A,B,C,D
3	3132	3317	185	EVLCPVAGADPELCVPGRMPTVLQCVNVSVVSEEVCS KLYDPLYHPSMFCAGGGQDQKDSCN	A,B,C,D
4	4588	4737	149	GDSGGPLICNGYLQGLVSFGKAPCGQVGVPGVYTNLC KFTIEWIEKTVQAS	A,B,C
5	12,536	12,607	71	ANLMTTAETPCTRLLGYYVHNVT	A,B,C,D

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 3. Predicted exons of the putative gene KLK-L2. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon prediction program ²
1	2383	2532	149	PNHKRRGLQPALAAAGTGHGKRIVLLGRPGASAVG AVSRTLFPFPEVSAE	B,D
2	2943	3205	262	SYTIGLGLHSLEADQEPGSQMVEASLSVRHPEYNRPLL ANDLMLIKLDESVSESDTIRSIASQCPTAGNSCLVSG WGLLANGE LT	A,B,C,D
3	3228	3413	185	EVLCPVAGADPELCVPGRMPTVLQCVNVSVVSEEVCS KLYDPLYHPSMFCAGGGHDQKDSCN	A,B,C,D
4	4686	4835	149	GDGGGLICNGYLQGLVSFGKAPCGQVGVPGVYTNL KFTEWIEK TVQAS	A,B,C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

607400 - 031499

Table 4. Predicted exons of the putative gene KLK-L3. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding sequence ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon prediction program ²
1	15,916	15,980	64	DYVIVAAECDVMDARICDRVTT	B,C
2	16,329	16,394	65	MGVSAELRACLAVTAVVRVVLG	A,B,C,D
3	17,904	18,163	259	HVLANNNDVSCDHPSNTVPSGSNQDLGAGAGEDARSDD SSSRIINGSDCDMHTQPWQAALLRPNQLYCGAVLVHP QWLLTAAHCRK	A,B,C,D
4	18,944	19,159	215	VYESGQQMFQGVKSIPHPGYSHPGHSNNLMLIKLNRR RPTKDVPRPINVSSHCPAGTKCLVSGWGTTKSPQ	C,D
5	19,245	19,378	133	HFPKVLQCLNISVLSQKRCEDAYPRQIDDTMFCAGDKA GRDSCQ	B,C
6	24,232	24,384	152	GDSGGPVVCNGSLQGLVSWGDYP CARPNRPGVYT NLC KFTKWIQETIQANS	A,B,C
7	25,286	25,410	124	LTFQEWKTDNKERDNWEAKAGESQVQEIE TILANMVK PPLY	B,C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 5. Predicted exons of the putative gene KLK-L4. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon prediction program ²
1	427	488	61	FQFPEAPQALVQEEKEEEQE	B,C
2	961	1038	77	LTMGRPRPRAAKTWMFLLLGGAWA	B,D
3	1878	2140	262	KYTVRLGDHSLQNKGPEQEIPVVQSIPHPCYNSSDVE DHNHDLMLLQLRDQASLGSKVKPISLADHCTQPGQKC TVSGWGTVTSPR	A,B,C
4	4252	4385	133	NFPDTLNCAEVKIFPQKKCEDAYPGQITDGMVCAGSSK GADTCQ	A,B,C,D
5	5922	6074	152	GDSGGPLVCDGALQGITSWGSDPCGRSDKPGVYTNICR YLDWIKK IIGSKG	A,B,C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail-2,
D = GENEID-3

Table 6. Predicted exons of the putative gene KLK-L5. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp) To(bp)		No. of bases	Translated protein sequence	Exon prediction program ²
1	70,473	70,584	111	VHFPTPINHRGGPMEEEGDGMA YHKEALDAGCTFQDP	A,B,C,D
2	70,764	70,962	189	CSSLTPLSLIPTPGHG WADTRAIGAECCRPNSQWPWQAGL FHLTRLFCGATLISDRWLLTAACRKK	A,B,C,D
3	73,422	73,687	265	LTSEACPSRYLWVRLGEHHLWKWEGPEQLFRVTDFPP HPGFNKDLSANDHNDIMLIRLPRQARLSPAVQPLNLS QTCVSPGMQCCLISGWGA VSSPK	A,B,C,D
4	76,305	76,441	136	LFPVTLQCANISILENLCHW AYPGHISDSMLCAGLWE GGRGSCQ	A,B,C,D

Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 7. Predicted-exons of the putative gene KLK-L6. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon predicti program
1	50,687	50,842	155	GDSGGPLVCGGVLQGLVSWGSVGPCGQDGIPGVY TYICKYVDWIRMMIRNN	C
2	53,550	53,607	57	AATAVSAATGPPEPQPQS	A,B
3	55,350	55,510	160	LVGGETRIIKGFECKPHSQPWQAALFEKTRLLCGA TLIAPRWLLTAAHCLKP	C,D
4	56,917	57,053	299	RYIVHLGQHNLQKEEGCEQTRTATESFPHPGFNNS LPNKHHRNDIMLVKMASPVSITWAVRPLTLSSRCV TAGTSCLISGWGSTSSPQCRCSTRGEPEGRG	A,B,C,
5	57,448	57,597	136	RLPHTLRCANITIEHQKCENAYPGNITDTMVCASV QEGGKDSCQ	A,B,C,
6	57,448	57,597	149	GDSGGPLVCNQLQGIISWGQDPCAJRKPGVYTK VCKYVDWIQE TMKNN	A,B,C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 8. Predicted exons of the putative gene KLK-L7. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon prediction program ²
1	25,460	25,728	268	GLKVVLGKHALGRVEAGEQVREVVHSIPHPEYRRSPTH LNHDHDIMLLELQSPVQLTGYIQTPLSHNNRLTPGTT RVSGWGTITSPQ	A,B,C,D
2	26,879	27,015	136	NYPKTLQCANIQLRSDEECRQVYPGKITDNMLCAGTKE GGKDS	A,B,C,D
3	28,778	28,963	185	GDGGGPLVCNRTLGYIVSWGDFPCGQPDRPGVYTRVS RYVLWIRETIRKYETQQQKWLKGPO	A,B,C
4	31,061	31,181	120	DSSGYSMGNRLRGPSEEGTALAQAKGDGSWTKVMAV EEAR	C,D

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 9. Predicted exons of the putative gene KLK-L8. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon prediction program ²
1	1588	1747	159	LSQAATPKIFNGTECGRNSQPWQVGLFEGTSLRCGGVL IDHRWVL TAAHCSG	A,B,C
2	3576	3851	275	VATGSRYWVRLGEHSLSQLDWTEQIRHSGFSVTHPGY LGASTSHEHDLRLRLRRLPVRVTSSVQPLPLPNDATA GTECHVSGWGITNHPR	A,B,C,D
3	4806	4939	133	PFPDLLQCLNLSIVSHATCHGVYPGRITSNMVCAGGVP GQDACQ	A,B,C,D
4	6023	6068	45	MSRALVLQGREFYSE	C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 10. Predicted exons of the unknown gene UG-1. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹		No. of bases	Translated protein sequence	Exon prediction program ²
	From(bp)	To(bp)			
1	43,117	43,194	77	AVEHKEAGTQSGNLQVPWPGWCGQA	A,C
2	43,433	43,607	174	RKQELMKHSSVMCQTQVSVNNQTQVSVNKMTKTPED AAFFEPSPFLEGKKTTELCCSHL	A,B,C
3	45,873	45,932	59	ALGIRSKFFCLALSGIIFR	C,D
4	47,004	47,111	107	EKSAVKKGTSQGDTDSKQRSWDSRHVAGPLQEEIK	A,C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis), C = Grail 2,
D = GENEID-3

Table 11. Predicted exons of the unknown gene UG-2. The translated protein sequences of each exon (open reading frame) are shown

Exon No.	Putative coding region ¹ From(bp)	To(bp)	No. of bases	Translated protein sequence	Exon prediction program ²
1	39,721	39,767	46	LTQPPEIHVQKSCNQ	A,B,C
2	44,129	44,641	512	PPLSLEPAVPERRTLRRNRRSLAALAPLTPDMLLLLPLL WGRERAEGQTSKLLTMQSSVTQEGLCVHVPCSFSYPS HGWITYPGPVVHGYWFREGANTDQDAPVATNNPARAV WEETRDRFHILLGDPHTKNCTLTSIRDARRSDAGRYFFRM EKGSIKWNYKHH RLSVNVT	B,C
3	44,843	45,121	278	LTHRPNILIPGTLESGCPQNLTCSPWACEQGTTPPMISWI GTSVSPLDPSTTRSSVLTIPQPQDHGTSLTQVTFPGAS VTTNKTIVHLNVS	A,B,C,D
4	45,327	45,374	47	PPQNLTMVFQGDGT	A,B,D
5	46,318	46,542	224	GQSLRLVCAVDAVDSNPPARLSLSWRGLTLCPSQPSNP GVLELPWVHLRDAAEFTCRAQNPLGSQQVYLNVSLQ	A,B,C
6	47,195	47,285	185	KATSGVTQGVVGGAGATALVFLSFCVIFV	A,B,C,D
7	49,136	49,322	185	GPLTEPWAEDSPPDQPPPASARSSVGEHELQYASLSFQ MVKPWDS RGQEATDTEYSEIKIHR	A,B,C

1. Nucleotide numbers refer to the related contig (see text and figure 1).

2. A = GeneBuilder (gene analysis), B = GeneBuilder (exon analysis); C = Grail 2,
D = GENEID-3

Table 12 . Homology between the predicted amino acid sequences of the newly identified putative genes and protein sequences deposited in GenBank

No.	Gene identity	Homologous known protein	Identity% (number of amino acids)
1	KLK-L1	• Human stratum corneum chymotryptic enzyme	47 (83/176)
		• Rat glandular kallikrein 10	45 (81/182)
		• Mouse glandular kallikrein K11	45 (77/172)
		• Human prostatic specific antigen	44 (74/168)
		• Human glandular kallikrein	44 (71/160)
2	KLK-L2	• Human stratum corneum chymotryptic enzyme	47 (104/220)
		• Rat trypsinogen V-B	47 (95/201)
		• Rat glandular kallikrein k2	45 (74/165)
		• Mouse glandular kallikrein K2	43 (97/223)
		• Human trypsinogen I	42 (95/227)
3	KLK-L3	• Human glandular kallikrein k2	57 (72/126)
		• Rat glandular kallikrein 10	55 (72/131)
		• Human trypsinogen I	54 (69/127)
		• Human stratum corneum chymotryptic enzyme	52 (74/141)
		• Rat glandular kallikrein k1	50 (69/137)
		• Human prostatic specific antigen	48 (84/174)
4	KLK-L4	• Rat trypsinogen II	47 (88/168)
		• Mouse trypsinogen	47 (93/200)
		• Human trypsinogen I	46 (87/188)
		• Human glandular kallikrein	46 (88/190)
		• Mouse glandular kallikrein K1	45 (87/194)
		• Human prostatic specific antigen	41(152/373)
5	KLK-L5	• Human glandular kallikrein	46(133/291)
		• Pig trypsin	44(134/302)
		• Human prostatic specific antigen	43(129/297)
		• Rat glandular kallikrein 8	42(130/307)
		• Rat trypsinogen II	42(131/279)
6	KLK-L6	• Rat glandular kallikrein k1	53(133/252)
		• Rat trypsinogen II	52(146/283)
		• Mouse trypsinogen	49(113/231)
		• Human trypsinogen II	47(139/297)
		• Human glandular kallikrein I	47(122/259)
		• Human prostatic specific antigen	47(116/249)
7	KLK-L7	• Rat trypsinogen III	54 (88/163)
		• Mouse trypsinogen	51 (95/188)
		• Mouse glandular kallikrein K1	48 (80/168)
		• Human trypsinogen I	45 (92/205)
		• Human prostatic specific antigen	45 (92/204)

No.	Gene identity	Homologous known protein	Identity% (number of amino-acids)
8	KLK-L8	<ul style="list-style-type: none">• Rat glandular kallikrein 7• Rat trypsinogen I• Pig trypsin• Human stratum corneum chymotryptic enzyme• Mouse glandular kallikrein K1	42 (57/136) 41 (56/138) 40 (55/139) 40 (52/131) 36 (49/135)
9	UG-1	<ul style="list-style-type: none">• Human zinc finger protein 91	43 (34/ 80)
10	UG-2	<ul style="list-style-type: none">• Human myeloid cell surface antigen CD33• Mouse myelin associated glycoprotein	59(128/216) 38 (53/141)

2025 RELEASE UNDER E.O. 14176

We Claim:

1. An isolated nucleic acid molecule which comprises:

- 5 (i) a nucleic acid sequence encoding a protein having substantial sequence identity preferably at least 70% sequence identity, with an amino acid sequence of **KLK-L1-KLK-L8** as shown in Tables 2 to 9;
- (ii) a nucleic acid sequence encoding a protein comprising with an amino acid sequence of **KLK-L1-KLK-L8** as shown in Tables 2 to 9;
- 10 (iii) nucleic acid sequences complementary to (i);
- (iv) a degenerate form of a nucleic acid sequence of (i);
- (v) a nucleic acid sequence capable of hybridizing under stringent conditions to a nucleic acid sequence in (i), (ii) or (iii);
- (vi) a nucleic acid sequence encoding a truncation, an analog, an allelic or species variation of a protein comprising with an amino acid sequence of **KLK-L1-KLK-L8** as shown in Tables 2 to 9; or
- 15 (vii) a fragment, or allelic or species variation of (i), (ii) or (iii).

60421260-331499

ABSTRACT OF THE DISCLOSURE

The invention relates to nucleic acid molecules, proteins encoded by such nucleic acid molecules; and use of the proteins and nucleic acid molecules

VERIFIED STATEMENT (DECLARATION) CLAIMING SMALL ENTITLEMENT STATUS (37 CFR 1.9(f) AND 1.27 (d)) - NONPROFIT ORGANIZATION			
			Docket No.
Serial No.	Filing Date	Patent No.	Issue Date
Applicant/ Eleftherios P. Diamandis Patentee:			
Invention: Novel Human Kallikrein-Like Genes			
<p>I hereby declare that I am an official empowered to act on behalf of the nonprofit organization identified below:</p> <p>NAME OF ORGANIZATION: <u>Mount Sinai Hospital</u> ADDRESS OF ORGANIZATION: <u>600 University Avenue</u> <u>Toronto, Ontario</u> <u>Canada</u> <u>M5G 1X5</u></p> <p>TYPE OF NONPROFIT ORGANIZATION:</p> <p><input type="checkbox"/> University or other Institute of Higher Education</p> <p><input type="checkbox"/> Tax Exempt under Internal Revenue Service Code (26 U.S.C. 501(a) and 501(c)(3))</p> <p><input type="checkbox"/> Nonprofit Scientific or Educational under Statute of State of The United States of America Name of State: _____ Citation of Statute: _____</p> <p><input checked="" type="checkbox"/> Would Qualify as Tax Exempt under Internal Revenue Service Code (26 U.S.C. 501(a) and 501(c)(3)) if Located in The United States of America Name of State: _____ Citation of Statute: _____</p> <p><input type="checkbox"/> Would Qualify as Nonprofit Scientific or Educational under Statute of State of The United States of America if Located in The United States of America Name of State: _____ Citation of Statute: _____</p> <p>I hereby declare that the above-identified nonprofit organization qualifies as a nonprofit organization as defined in 37 C.F.R. 1.9(e) for purposes of paying reduced fees to the United States Patent and Trademark Office regarding the invention described in:</p> <p><input checked="" type="checkbox"/> the specification to be filed herewith.</p> <p><input type="checkbox"/> the application identified above.</p> <p><input type="checkbox"/> the patent identified above.</p> <p>I hereby declare that rights under contract or law have been conveyed to and remain with the nonprofit organization with regard to the above identified invention.</p> <p>If the rights held by the above-identified nonprofit organization are not exclusive, each individual, concern or organization having rights to the invention is listed on the next page and no rights to the invention are held by any person, other than the inventor, who could not qualify as an independent inventor under 37 CFR 1.9(c) or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d) or a nonprofit organization under 37 CFR 1.9(e).</p>			

Each person, concern or organization to which I have assigned, granted, conveyed, or licensed or am under an obligation under contract or law to assign, grant, convey, or license any rights in the invention is listed below:

- no such person, concern or organization exists.
- each such person, concern or organization is listed below.

FULL NAME _____

ADDRESS _____

Individual Small Business Concern Nonprofit Organization

FULL NAME _____

ADDRESS _____

Individual Small Business Concern Nonprofit Organization

FULL NAME _____

ADDRESS _____

Individual Small Business Concern Nonprofit Organization

FULL NAME _____

ADDRESS _____

Individual Small Business Concern Nonprofit Organization

FULL NAME _____

ADDRESS _____

Individual Small Business Concern Nonprofit Organization

Separate verified statements are required from each named person, concern or organization having rights to the invention averring to their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is no longer appropriate. (37 CFR 1.28(b))

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application, any patent issuing thereon, or any patent to which this verified statement is directed.

NAME OF PERSON SIGNING: Terry Donaghue

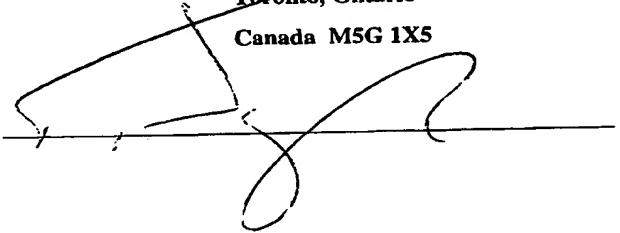
TITLE IN ORGANIZATION: Director, Technology Transfer & Industrial Liaison

ADDRESS OF PERSON SIGNING: Mount Sinai Hospital

600 University Avenue

Toronto, Ontario

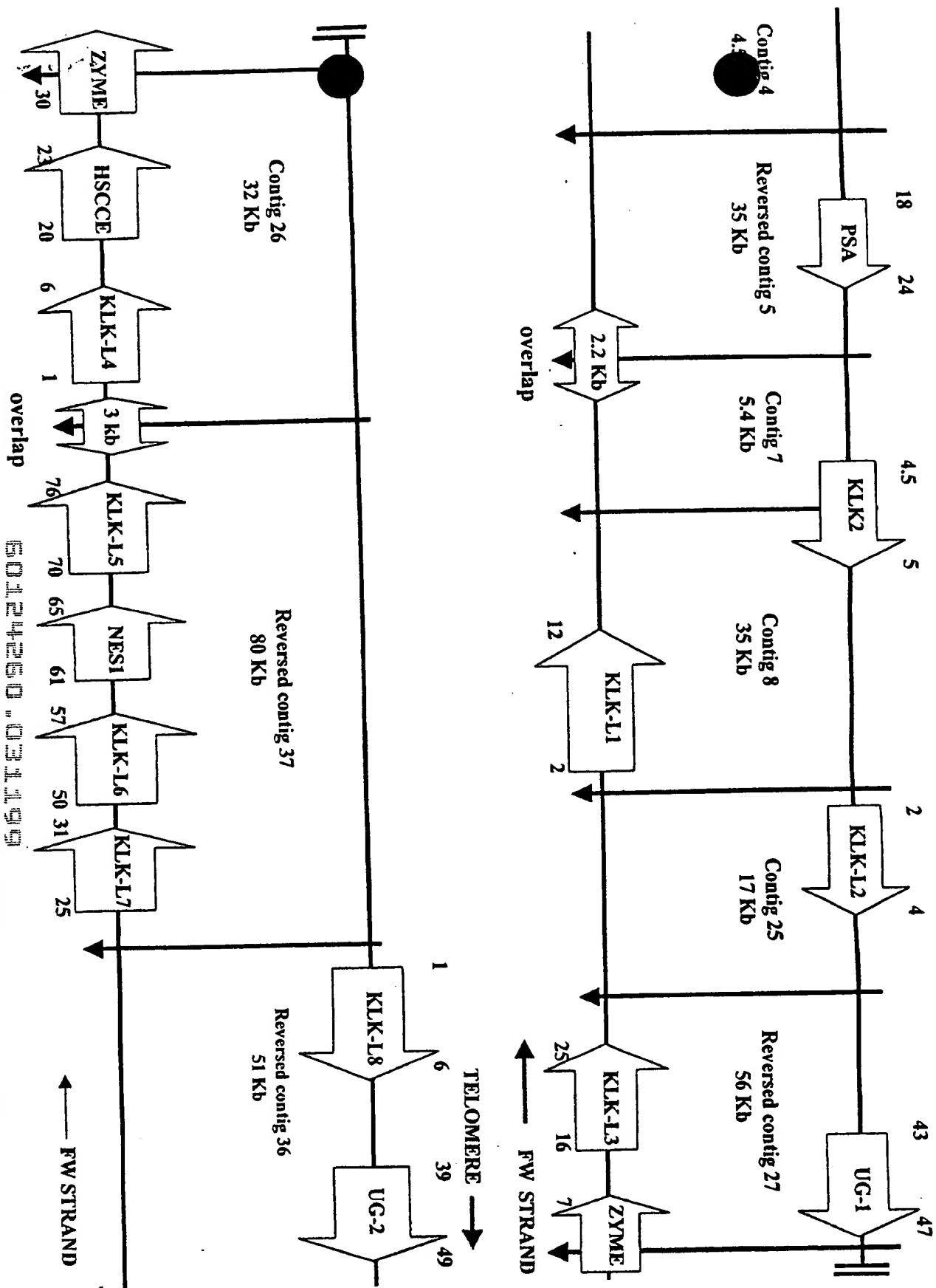
Canada M5G 1X5

SIGNATURE: 

DATE: March 10, 1999

FIGURE 1

↑ CENIROMEKE



THIS PAGE BLANK (USPTO)